DOCUMENT RESUME

ED 354 614 EA 024 677

AUTHOR Martinez, Carlos
TITLE Policy and Technical Issues on the Identification of Schools for Improvement of Chapter 1 Programs.
PUB DATE Apr 91
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Accountability; *Achievement Rating; *Compensatory Education; *Educational Assessment; *Educational Improvement; *Educationally Disadvantaged; Elementary Secondary Education; *Evaluation Criteria; Instructional Effectiveness; Public Schools; School Effectiveness
IDENTIFIERS Education Consolidation Improvement Act Chapter 1; *Hawkins Stafford Act 1988

ABSTRACT

Identification of schools for improvement of Chapter 1 Programs raises the challenge of selecting the best technical procedures meeting the law and usefully guiding improvement efforts. Failure to demonstrate gains in student achievement has been operationalized on two levels: aggregate performance and desired outcomes. Aggregate performance evaluations have typically relied on norm-referenced testing methods. Though dependency on norm-referenced test data is inadvisable, few school districts have used the desired-outcomes approach, which relies on criterion-referenced tests, state assessment, and end of unit tests, among others. These districts fear that such an approach exposes them to an additional layer of entrapment. A method of triangulation or composite analysis could make a single determination of the need for improvement based on multiple data sources, so that the desired-outcomes method is seen as a contributor rather than as entrapment. Identifying a need for program improvement will guide reform efforts. Methods are needed to ensure teacher confidence in the measures used to gauge progress so that the hope for institutional change will not be limited. (TEJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED354614

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Martinez, Jr.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Policy and Technical Issues on the Identification of Schools for Improvement of Chapter 1 Programs

Carlos Martínez
Compensatory Education Programs
Office of Elementary and Secondary Education
United States Department of Education

Presentation at the Annual Meeting of the
American Educational Research Association
Chicago, Illinois
April 1991

This paper is intended to promote the exchange of ideas among researchers and policy makers. The views are those of the author, and no official support by the U.S. Department of Education is intended or should be inferred.

EA 024 677

BEST COPY AVAILABLE

The Hawkins-Stafford Amendments of 1988 (P.L. 100-297) has caused the education of disadvantaged students to be viewed in a new light. Chapter 1 of Title I of the Hawkins-Stafford Elementary and Secondary School Improvement Amendments (which amends the Elementary and Secondary Education Act of 1965) marks new thinking in compensatory education that emphasizes advanced as well as basic skills, school level accountability, and parental involvement. These legislative directions reflect educators' understanding that students learn within the context of a school and a home. The challenge facing evaluators in this era of school and program reform is to implement the best technical procedure that meets the requirement of the law and provide the most useful information to guide improvement efforts.

The program improvement language in the Hawkins-Stafford Amendments (Sections 1020 and 1021) is the driving force behind the reform initiative in Chapter 1. It intends to identify programs operating in schools whose students fail to show gains in achievement. Failure to demonstrate gains has been operationalized on two levels: aggregate performance and desired outcomes. This presentation will focus on the procedures used to identify programs in schools that need improving based on aggregate performance and progress toward reaching desired outcomes.

Aggregate Performance

Aggregate performance refers to a review of Chapter 1 students' test results using the school they attend as a unit of analysis. The analysis and instrumentation that can be used is specified in Section 1019, which describes the evaluation requirement, and Section 1435, which addresses the national standards for local evaluation. The law allows the use of the Title I Evaluation and Reporting System (TIERS) as a model for the national standards. The national standards refer to minimal specifications of the quality of data that is necessary for national aggregation. It does not refer to standards of student or program performance.

TIERS advances three evaluation models for local programs to select as the method for evaluating and reporting the effectiveness of their local Title I (Chapter 1) program. The model that was almost exclusively selected was Model A, a design that relied on norm-referenced testing and a common reporting scale of normal curve equivalents (NCEs). This model was intended to produce achievement information that could be aggregated across school districts and States. It was developed before there was a statutory requirement to review school level performance.

The local evaluation requirement (Section 1019) and the national evaluation standards have been regulated in § 200.80 to mean that all Chapter 1 participants in grades 2

through 12 must be tested on a standardized norm-referenced test in the basic and more advanced skills in all the subject areas in which the student receives assistance through Chapter 1 (usually reading, mathematics, and language arts). A district may choose to test only in the advanced skills, as reported on a reading comprehension or mathematics applications subtests. Language arts may be evaluated using a reading test. Students are to be tested annually using a fall to fall or spring to spring testing cycle. Exceptions to this requirement are students participating in programs designed to teach limited English proficient students and students below the second grade.

To review school level performance, a school districts must aggregate the test results for all Chapter 1 students with matched scores across all grades served and determine if they have made gains over a year's time. A district may choose to use either the mean or the median to make a determination of aggregate gain. This option was offered to help schools with small Ns contend with the effects of extreme scores. This choice will apply for all schools within the district, however.

Desired Outcomes

The Department does not wish for districts to make a determination on program improvement based on solely one measure. To avoid this, statutes and regulations require an annual review of a school's progress toward meeting the program's desired

outcomes as stated in their application to the State for Chapter 1 funds. Desired outcomes must be stated in measurable terms and must apply to all students in the program. Substantial progress toward meeting the desired outcomes must also be stated and for multi-year applications it can be stated incrementally over time.

The data that can be used to determine substantial progress towards desired outcomes is varied and is bound only by an administrator's willingness to accept these measures. Among the possibilities are criterion-referenced test, State assessment tests, end of unit tests, classroom grades, observation checklists, and new techniques in performance and portfolio assessments. Regulations also reguire districts to use the norm-referenced aggregate performance standard as a minimal desired outcome.

Unfortunately, few districts have developed desired outcomes that are used to determine school level performance and have relied on norm-referenced testing as the only indicator. The reason is that desired outcomes places Chapter 1 programs in what some program administrators call "double jeopardy." Desired outcomes do not replace aggregate performance as indicators for program improvement, but must be considered along with the results of norm-referenced testing. This creates a second layer of entrapment that administrators would like to avoid.

Identification of Schools

The primary problem in identifying schools with Chapter 1 programs that need improvement is that it labels a school as being unsuccessful in teaching disadvantaged students. District and school administrators are driven to avoid this label since it reflects on the school as a whole. As it stands now, a school with a Chapter 1 program providing services in all three subject areas (reading, mathematics, and language arts) will be assessed under five criteria for program improvement. Including desired outcomes will increase the probability of identification by adding even more criteria. Administrators who wish to diminish the probability of being identified will take a minimalist perspective on assessment and choose to test only in reading comprehension and mathematic applications. Though this narrows the scope of evaluation, it also streamlines the burden of testing, which is also appealing to program administrators.

Another issue in the identification of schools for program improvement is the treatment of schools with small Ns. Permitting the use of the median instead of the mean has limited appeal since this option must be applied to all schools in the district, not just those with small Ns. It does not address the underlying technical issue of a widening confidence band as the N decreases. Regulations exempt programs with less than 10 students from identifying schools for program improvement, but does not exempt a

school whose number of students with matched scores is less than 10.

These two issues lead us to consider two policy questions: How can school districts be encouraged to use desired outcomes and how can program efficacy be determined if it cannot be adequately evaluated with a norm-referenced test (whether due to small Ns, student migration or other factors that may introduce error)?

The question of engaging districts in exploring desired outcomes as a measure to determine the need for program improvement is best solved by removing the "double jeopardy" threat. A technical contribution would be to develop a method of triangulation that would make a single determination of need for program improvement based on multiple data sources. A triangulated or composite analysis would weigh all the outcome measures but not depend on any one measure to make a determination. Once desired outcomes are seen as contributors rather than triggers for the determination for improvement, program administrators may be more willing to explore their possibilities.

There are many issues that challenge the adequacy of norm-referenced testing as the basis to determine aggregate performance and plan for school level program improvement. It is doubtful that they can all be addressed here. But policy makers can contribute to the solution of the small N issue by requiring school that cannot

produce 35 matched scores to determine the need for program improvement on the basis of desired outcomes only. Schools reviewing aggregate performance based norm-referenced test data with such small Ns would do better to examine other desired outcomes than to analyze data of questionable validity.

The Status of Program Improvement

Program improvement is in its second year of implementation. Even with all the concerns discussed here and the very cautious posture taken by program administrators 6,329 schools with Chapter 1 programs have been found to be in need of program improvement (MacDonald, 1991:28). This represents almost 12 percent of 53,000 schools operating Chapter 1 programs.

Once a school is identified for program improvement the usual step taken immediately is to seek technical assistance in verifying the quality of achievement data. After the data are verified a process of disaggregation is conducted for each subject area, grade level or even classroom to determine where program improvement activities should be focused. Planning for improvement should consider these data, but are not limited to them. Implementation of the plan should commence as soon as possible. If the school shows progress over a year during the three year plan, then the school does not need to continue implementation. If the school fails to demonstrate

improvement after three years, it must collaborate with the State education agency in developing a new plan. Next year will mark the point where some school districts will enter planning with their States.

Conclusion

Dependency on norm-referenced test data to identify and plan for school improvement is not advisable. Technical considerations alone require the use of caution in interpreting these results. The Department has responded to these concerns by permitting and encouraging the use of other measures in making these determinations. Unfortunately, these options increase the risk of identification that carries a label that school and program administrators wish to avoid. Solutions to these problems require contributions from both policy and technical experts. But there is a greater question that remains, that is, whether test data that is used primarily to sort, rank, and select participants should also be used to identify schools with programs that need improving and whether that data should guide school reform efforts. Any change in teaching practice requires that the stakeholders have confidence in the measures used to gauge progress. Without that confidence any hope of institutional change would be limited. Practice will be guided by the narrow scope of vision that characterizes traditional testing. A broader vision of teaching and program operation must be accommodated by assessment procedures in order for program improvement to flourish.

References

Congress of the United States (1988)
The Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvements Amendments of 1988 (Public Law 100-297, 20 U.S.C. 2835).

MacDonald, J. T. (1991)
Annual Report on Program Improvement (Memorandum). Washington, DC: U.S. Department of Education.

United States Department of Education (1989)
Chapter 1 Program in Local Education Agencies; Final Regulations (34 CFR Part 75 et al.). Washington, DC: Author.

United States Department of Education (1990)
Chapter 1 Policy Manual. Washington, DC: Author.